

Meetings of the Metadata Reviewers Community

(most recent meeting first, then reverse chronological to oldest meeting last)

Mon. December 5, 2016, 2pm - 3pm EDT

Proposed agenda:

1. Burning questions? Metadata nightmares? Brilliance to brag about?
2. Look at suggested revision to the USGS Data Management website > Publish/Share > Data Release <<https://www2.usgs.gov/datamanagement/share/datarelease.php>> > Section 5.
3. Are we ready to start reviewing the [data](#) and [metadata review](#) checklists? (Or wait until January?)
4. Do we want to sponsor training at the CDI Workshop?
5. Next meeting (not Jan. 2).

Notes from meeting:

1. Colin reports seeing good metadata at FORT. Bill is working with data that will be contributed to the California environmental data repository, and reports that our standards for metadata and review are more stringent than theirs.
There has been some recent email discussion of where data producers' ORCID IDs should go in the metadata record. There seems to be no place to put ORCID IDs where they would be immediately useful in systems like the Science Data Catalog or data.gov, but there are several places where they might reasonably be found and would not cause a record to fail standard validation checks. Peter Schweitzer will start a discussion at our community confluence site so that we can decide on a consistent approach.
2. Alan introduced the revision to section 5 of the website, explaining that much of the information in the present section is off subject, for example, about publications that are not data. Fran added that the new USGS policy is that no data is interpretive, so we decided to drop the sentence about interpretive data from the revised section. We would like to add specifics of how IPDS is used for data and metadata reviews, and Fran made a suggestion for that in the Google document. We would also like the webpage to provide more easily found links to guidance and policy.
Susie showed the IP record in progress for a Santa Cruz data release in ScienceBase. The record shows original metadata files and reviewed metadata files, as well as reviewed ScienceBase pages. This case does not have the metadata harvested from ScienceBase to Science Data Catalog. Conversation continued on the question of whether the short metadata records for ScienceBase project pages, which do not include data but provide a description of a collection of data, need to be compliant with metadata validation, for example, by mp. Tamar shared information that in the future such metadata will be harvested and thus will need to be validated. Peter said that a basic metadata record that only has sections 1 & 7 could be validated. ISO metadata more intrinsically accounts for relationships between collections and the items they contain.
Decisions: We will leave the revision on Google docs and encourage community members to suggest improvements, using "suggesting" mode instead of "editing" (the mode choice is available in the upper right corner, under the Comments button). Also suggest guidance and policy links that should be provided on the webpage. Fran will negotiate the webpage changes with Viv Hutchison.
3. Peter will put the data and metadata review checklists on Google docs so that community members can start suggesting modifications (see links below). Our goal is to have fairly generic checklists, helpfully grouped and chunked, with links to more detailed lists for particular kinds of data.
4. We did not have time for discussion of the CDI workshop.
5. We decided to skip the January phone call, since Jan. 2 is a holiday and the Data Management Working Group is likely to be meeting on Jan. 9.

Other discussion topics:

Briefly raised, what about data that is included in administrative reports and proprietary data?

Briefly raised, how can we deal with the issue of links to files changing in ScienceBase, when the data is modified, and the challenge of keeping links correct in metadata?

Google docs for community review before our Feb. meeting:

[Data Review Checklist](#) is a copy of the existing checklist formatted as a Google Docs and shared for edit and comment.

[Guidelines for Metadata Review of Data](#) is a copy of the existing checklist formatted as a Google Docs and shared for edit and comment.

Mon. November 7, 2016, 2pm - 3pm EDT

Proposed agenda:

1. Burning questions? Metadata nightmares? Brilliance to brag about?
2. Start reviewing the [data](#) and [metadata review](#) checklists on the [Data Management Website](#).
3. Any ideas about how we might get together at the CDI Workshop?
4. Next steps?

Notes from meeting:

Metadata Reviewers Community
Meeting: 20161107

Peter Schweitzer leading, in Fran's absence, with input from Alan Allwardt, VeeAnn Cross, and the group
Notes by Alan Allwardt

Agenda Item 1. Burning questions

Peter Schweitzer: told story of someone asking him what to do about a non-geospatial dataset for which the metadata failed mp because there was no spatial domain information. In the past Peter would have recommended ignoring the mp error, but now he recommends entering a global extent to avoid validation errors in downstream catalogs like [data.gov](#).

Lisa Zolly: confirmed that [data.gov](#) will flag and quarantine CSDGM records lacking a spatial domain (USGS Science Data Catalog will not).

Members of the group shared their strategies in dealing with metadata for non-spatial data: some create global spatial extents; others will use the bounding box of the parent project for non-geospatial, supplementary or lab data. It was generally agreed that using the coordinates for the science center where non-geospatial lab results were obtained is a BAD idea.

ACTION ITEM: Peter will add a paragraph to his "Substantive review of metadata" training page <<http://geo-nsdi.er.usgs.gov/validation/how-to-review/elements.html>> to deal with spatial domain conundrums.

Agenda Item 2. Revising the data review and metadata review checklists

Peter suggested stepping back from the checklists and look at the context in which they are presented: USGS Data Management website > Publish/Share > Data Release <<https://www2.usgs.gov/datamanagement/share/datarelease.php>> > Section 5.

ACTION ITEM: After extensive discussion, the group decided that the text of Section 5 -- which provides context for the checklists -- should be revisited and revised as necessary FIRST, and only then should we consider how to revise the checklists themselves. (Revising the text of Section 5 will inform the process of revising the checklists.) This plan met with general approval. Alan will begin revising Section 5 and get input from Peter, VeeAnn, and Fran before it is posted on Google Docs for the group to consider.

Highlights of the discussion leading to the action item above:

Peter: data review and metadata review not clearly separated (lots of agreement on that point from the group).

VeeAnn: noted that the revision dates of the checklists (March/April 2014) predate the OSQI IM on data management, data release and metadata (IM 2015-01 through 2015-04): <<https://www2.usgs.gov/usgs-manual/95imlist.html>>. We need to examine the checklists and, at the very least, bring them in alignment with these IM. NOTE: IM OSQI 2015-03, Section 5A <<https://www2.usgs.gov/usgs-manual/im/IM-OSQI-2015-03.html>> links directly to the checklists, so we are constrained to revising the checklists individually (we can't combine them, for instance).

Several members of the group shared how they've used the data review and metadata review checklists in their science centers: they've used the checklists as a starting point for creating more specific guidance documents for their particular science centers. Alan created a thread in the Metadata Reviewers Forum where members can share their experiences in adapting the checklists (with encouragement to upload examples of specialized checklists, review templates, etc.): <<https://my.usgs.gov/confluence/pages/viewpage.action?pageId=558860218>>.

Peter created another thread in the Forum for members to share their thoughts on how the data/metadata review process might be documented for IPDS: <<https://my.usgs.gov/confluence/pages/viewpage.action?pageId=558860180>>.

Peter: What about revising "Metadata in Plain Language" <<http://geology.usgs.gov/tools/metadata/tools/doc/ctc/>> so that it is less CSDGM-specific?

VeeAnn: noted that two reviews are necessary -- of data and metadata -- although they can be performed by the same person. She proposed another strategy: use two people. The first would emphasize the data review (but also look at the metadata), the second would emphasize the metadata review (but also look at the data).

Agenda Item 3. 2017 CDI Workshop

Brief discussion at the top of the hour, will continue next time.

Peter suggested considering hands-on training, in one of the following areas:

- Helping metadata reviewers who are new to the USGS
- Strategies for documenting the review process
- Keywords (utilizing controlled vocabularies)
- Strategies for integrating data and metadata reviews
- Sharing useful tricks of the trade

Mon. October 3, 2016, 2pm - 3pm EDT

Proposed agenda:

1. Burning questions? Metadata nightmares? Brilliance to brag about?
2. [Keywords in Metadata](#), a presentation from the USGS Thesaurus Team
3. Next steps?

Notes from meeting:

Question: How can we deal with metadata records that use the EML standard?

- Lisa Zolly: EML metadata will need to be converted to the CSDGM standard, or in the future to ISO. There is a XSL transform to do that; [email Lisa](#) directly if you need it.
- Peter Schweitzer: If you're worried about losing information in the format conversion, you can link to the original EML record.

Question: Metadata records are being written by project members who are not USGS employees but are students at a university, so they are unable to authenticate with OME which only uses USGS active directory credentials. Can we get them guest user permission?

- Lisa Zolly: The OME team intends to do that some day, but does not have the resources to do it any time soon. OME relies on Active Directory for authentication; a separate module could be leveraged for external accounts, but the database would need to be built for it, and CSASL would have to dedicate staff to supporting management of non-AD accounts.
- Tom Burley: Could the students sign up as USGS volunteers? (Well, no, that is expensive.) Or use Metavist.
- Aaron Freeman: Could the students use Metadata Wizard in the Esri context? The record could be exported and more information could be added in Metavist.
- Follow-up: Isn't Metadata Wizard being de-coupled from the Esri environment? (That's the hope, but no resources to do it yet. You could import a CSV into a geodatabase, though.)
- NOAA has shut down Mermaid, and EPA has also shut down its CSDGM metadata editor, because both are going to the ISO metadata standard.

Presentation, see Peter Schweitzer's outline linked in the agenda above.

- Peter's concern is how well metadata works in situations where people are using it to find information. Because people often don't know what to ask for, what to call it, or who to ask, Peter was led to the use of controlled vocabularies.
- He prefers keywords that say what the data are, instead of those that say what purposes the data could be used for.
- He cautions that names can be interpreted in different ways, so more keywords are necessary to clarify what the data are.

Lisa Zolly shared the [list of USGS Thesaurus terms that is being used in the USGS Science Data Catalog](#) to provide a browse interface. SDC also allows full-text searches of metadata records, with some fields being weighted more heavily than others. As more metadata records provide one of the keywords on the browse list, the interface will be quicker and better.

General tips for metadata reviewers:

- Some accurately spelled keywords from accurately identified thesauri are an important part of good metadata.
- Metadata keywords should include some general terms that allow people to narrow down their search, and also specific terms that allow people to rule out the data sets that are not what they need, and rule in data sets that might be what they need.
- USGS has some tools to help reviewers compare keywords to thesauri, and more thesauri could be added to them.
- If you want your data to show up well in the USGS Science Data Catalog, make sure there is a keyword from Lisa's list.

First meeting, Mon, September 12, 3pm – 4pm EDT (after the CDI Data Management Working Group meeting).

Proposed agenda:

1. Burning questions? Metadata nightmares? Brilliance to brag about?
2. Our Community
 - a. Focus: Review of USGS metadata
 - b. Community: Share knowledge, questions, and puzzles

- c. Knowledge: Develop, share, and maintain know-how for review of USGS metadata
- 3. Community Resources
 - a. Confluence Site: Member list, link to training, examples, discussion “forum”
 - b. Data Management Website
- 4. Next steps?
 - a. A session on keywords?
 - b. Distinguish clear USGS requirements from matters for criteria and considerations?
 - c. Help desk? Monthly meetings? Review & revise online checklist?

Notes from meeting:

Question: Will there be a similar group for data review, or does this group include data review?

- We agreed to expand our scope to include data review, especially technical aspects of data release such as packaging, format, and documentation.
- We agreed that a good metadata review requires looking at the data to ensure that the metadata represents it correctly.
- USGS policy requires two reviews (metadata and data) and not two separate reviewers, but this is a minimum standard. Science center directors can require additional reviews before they approve the data release. Our group agreed that two different reviewers looking at a data release would be a good thing for ensuring quality, and for high-profile datasets more than two might be good – we might want to think about defining levels of review. The Alaska Science Center has a scientist peer review the data content and a data manager review technical aspects including metadata.

Observation: As USGS implements the new policy with unprepared reviewers, it's almost inevitable that some “horror story” data will be released that will be embarrassing. It would be good for us to keep our ears to the ground – who needs help in reviewing data and metadata?

Issue: Metadata writing and reviewing are a significant time investment. How can we help our scientists and managers plan realistically?

- We could “pass around notes” about how long it takes, producing a community estimate that could be shared more widely.
- Data management plans will be helpful when they are required.
- Metadata reviewers tend to also become metadata counselors, helping new metadata writers avoid difficult and time-consuming approaches, and even providing training.
- Another way of helping research projects get started with metadata is to provide templates customized with appropriate contacts and disclaimers, which simplifies the project's work, helps standardize their metadata, and makes review easier.
- We agreed to enlarge the scope of our community to include metadata counseling, training, and resources.

Future community meetings.

- We agreed to meet monthly at 2:00 Eastern Time on the first or third Mondays of the month. (This is the same time of the week as the CDI Data Management Working Group and the Science Center Points of Contact for the new policies, but different weeks.) If we meet on the first week, the third week might be used for subgroup meetings.
- We agreed to have a session on keywords.
- We can post possible discussion topics on the confluence forum.

Future community activities.

- Similar to the library of recommended disclaimers, we could recommend wording that can be used in metadata records for referring to information that is documented somewhere else, for example in data dictionaries, “techniques and methods” publications, or NWIS documentation.
- We could revise the checklists for data reviewers and metadata reviewers on the USGS, incorporating the separate list that VeeAnn and Peter provided as part of their training.
- We will start a confluence forum topic on recommended resources, and start it off with the “green workbook” which several people recommend.

A question was raised about metadata for a geodatabase that includes multiple data sets. The discussion was diverted to one about acceptable data release formats for GIS data. SDTS has been withdrawn with no replacement, geodatabases are proprietary, shapefiles are said to have problems with spatial reproducibility. The discussion will continue on the forum page on our confluence site. The larger question is how we as reviewers should advise authors about data distribution packaging (convenience, clarity, longevity).